Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [F-18]FDG-PET/CT Studies

van Velden, Floris H. P.; Kramer, Gerbrand M.; Frings, Virginie; Nissen, Ida A.; Mulder, Emma R.; de Langen, Adrianus J.; Hoekstra, Otto S.; Smit, Egbert F.; Boellaard, Ronald

Published in:
Molecular Imaging and Biology

DOI:
10.1007/s11307-016-0940-2

MOLECULAR IMAGING AND BIOLOGY

Mol Imaging Biol (2016) 18:788–795
DOI: 10.1007/s11307-016-0940-2

Published Online: 26 February 2016

RESEARCH ARTICLE

Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [^{18}F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation

Floris H. P. van Velden,[1,2] Gerbrand M. Kramer,[1] Virginie Frings,[1] Ida A. Nissen,[1] Emma R. Mulder,[1] Adrianus J. de Langen,[3] Otto S. Hoekstra,[1] Egbert F. Smit,[3,4] Ronald Boellaard[1,5]

[1]*Department of Radiology and Nuclear Medicine, VU University Medical Center, PO Box 70571007MB, Amsterdam, The Netherlands*
[2]*Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands*
[3]*Department of Pulmonary Diseases, VU University Medical Center, Amsterdam, The Netherlands*
[4]*Department of Thoracic Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands*
[5]*Department of Nuclear Medicine and Molecular Imaging, University Medical Center Groningen, Groningen, The Netherlands*

Abstract

Purpose: To assess (1) the repeatability and (2) the impact of reconstruction methods and delineation on the repeatability of 105 radiomic features in non-small-cell lung cancer (NSCLC) 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) positron emission tomorgraphy/computed tomography (PET/CT) studies.

Procedures: Eleven NSCLC patients received two baseline whole-body PET/CT scans. Each scan was reconstructed twice, once using the point spread function (PSF) and once complying with the European Association for Nuclear Medicine (EANM) guidelines for tumor PET imaging. Volumes of interest ($n = 19$) were delineated twice, once on PET and once on CT images.

Results: Sixty-three features showed an intraclass correlation coefficient ≥ 0.90 independent of delineation or reconstruction. More features were sensitive to a change in delineation than to a change in reconstruction (25 and 3 features, respectively).

Conclusions: The majority of features in NSCLC [^{18}F]FDG-PET/CT studies show a high level of repeatability that is similar or better compared to simple standardized uptake value measures.

Key words: PET/CT, Repeatability, Radiomics, Tracer uptake heterogeneity, Non-small-cell lung cancer (NSCLC)

Introduction

Lung cancer is the leading cause of cancer death, with more than 1.6 million deaths worldwide in 2012 [1]. About 80–85 % of the cases are classified as non-small-cell lung cancer (NSCLC) [2]. Early assessment of response to treatment (e.g., radiotherapy and/or chemotherapy) is essential to

Electronic supplementary material The online version of this article (doi:10.1007/s11307-016-0940-2) contains supplementary material, which is available to authorized users.

Correspondence to: Floris van Velden; *e-mail:* f.h.p.van_velden@lumc.nl

Table 1. Patient demographics

Parameter	Value
Gender	
Male	7
Female	4
Age (year)	
Median	61
Range	45–66
Weight (kg)	
Median	76
Range	57–114
Stage	
III B	4
IV	7
Histology	
Adenocarcinoma	8
Squamous cell carcinoma	3
Type of lesion	
Primary	6
Metastasis	13
Localization	
Lung	5
Mediastinum	8
Hilum	2
Clavicular region	4
Lesion volume (ml)	
Median	39
Range	18–702

determine which patients will benefit from treatment and which may require treatment adaptations, paving the way for personalized cancer therapies [3]. Several studies have demonstrated the potential of positron emission tomography/computed tomography (PET/CT) to assess the effects of treatment for NSCLC patients early using 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) [4–7]. Although the benefits of new response metrics such as the metabolically active tumor volume (MATV) and total lesion glycolysis (TLG) are currently under investigation, response to treatment is predominantly measured using the maximum standardized uptake value (SUV_{max}) obtained within a tumor [8]. However, SUV_{max} is not capable to capture all forms of responses accurately. For instance, SUV_{max} can only measure response accurately if there is a global change in tracer uptake, i.e., in absence of a spatially heterogeneous response [9]. In addition, since SUV_{max} only involves a single voxel, it is inherently unable to capture intratumor heterogeneity and unable to measure a change in the shape or volume of (the metabolically active part of) the tumor. In recent years, various advanced quantitative imaging features, so-called radiomic features, have been proposed and investigated for their potential to quantify tracer uptake, tracer uptake heterogeneity, and/or (metabolically active) tumor geometry [9–21]. The term radiomics refers to studies that extract a large amount of advanced quantitative imaging features from medical imaging studies, e.g., PET/CT studies, as a basis for characterizing a specific aspect of patient health [22–24].

Several challenges have been identified that need to be addressed before radiomic features can be used in clinical practice, including the standardization and robustness of selected features [21]. For standardization, it is of utmost importance to identify which radiomic features are sensitive to a change in reconstruction settings [25–27] or to a change in delineation [11, 26]. For instance, radiomic features that can characterize tracer uptake heterogeneity may treat both partial volume effects and noise as heterogeneity [9]. Although it has been shown that several radiomic features are not sensitive to partial volume effects and noise when extracted from PET/CT response data of esophageal carcinoma patients [26], it has been shown that some features do require image denoising and partial volume correction prior to extraction [9, 26]. Recently, two studies [25, 27] investigated the effects of different reconstruction settings on the values obtained from various texture-based features and indicated a need for standardization. Note that for response monitoring studies, it is important to know whether an observed change in tracer uptake, tumor geometry, or tracer uptake heterogeneity is due to a true response or methodological variation (i.e., biological, technical, or observer variability). Therefore, it is essential to assess the repeatability of these radiomic features. However, to the best of our knowledge, the effects of reconstruction and delineation on the repeatability of a large set of radiomic features, including intensity-, shape-, and texture-based features, have not yet been assessed.

Therefore, the aim of this study was to assess the repeatability of various radiomic features in NSCLC [^{18}F]FDG-PET/CT studies, taking different reconstruction settings and delineation methods into account. To assess the impact of different reconstruction settings, PET data were reconstructed twice using settings that either ensure harmonization (i.e., complied with the European Association of Nuclear Medicine (EANM) guidelines for tumor PET imaging [28]) or are more state of the art (i.e., use of a resolution model during image reconstruction). To assess the impact of delineation, volumes of interest (VOIs) were defined manually on (low-dose) CT images and semi-automatically on PET images. CT-based delineation was explored to illustrate the effects when using the anatomical volume of a tumor, thereby potentially capturing a higher level of tracer uptake heterogeneity within a VOI (e.g., by the inclusion of necrotic areas) compared to semi-automatic threshold-based isocontour methods on PET data.

Materials and Methods

Patient Data

Eleven NSCLC patients (Table 1) received double-baseline whole-body [^{18}F]FDG scans that were acquired on a time-of-flight (TOF) PET/CT scanner (Philips Healthcare, Cleveland, OH). The time interval between first and second baseline scans was less than 3 days (1.3 ± 0.5 days). This prospective study has been approved by the institutional review board and is part of a study that has been registered in the Dutch trial register (www.trialregister.nl;

Table 2. Implemented radiomic features with corresponding literature references describing the features

Group	No. of features	Names of radiomic features	Described in
Intensity	27	Maximum standardized uptake value (SUV_{max}), mean SUV (SUV_{mean}), mean SUV of a sphere of 12-mm diameter (SUV_{peak}), coefficient of variation (COV), total lesion glycolysis (TLG), mean SUV of maximum SUV and the six adjacent voxels (SUV_{star}), minimum SUV (SUV_{min}), range of SUV (SUV_{range}), median SUV (SUV_{median}), standard deviation (SD), skewness, kurtosis, mean absolute deviation, median absolute deviation, mean Laplacian, total energy, variance, root-mean-square (RMS), Moran's I, Geary's C, uniformity[a], entropy[a], local entropy[a], and area under a cumulative (AUC) SUV-volume histogram	[9–14]
Shape	9	Compactness A, compactness B, sphericity, disproportion, surface area, metabolically active tumor volume (MATV) or anatomical volume (AV), surface to volume ratio (S2V), surface of an equivolumetric sphere to volume ratio ($S2V_{eq}$), and radius of an equivolumetric sphere	[11]
Texture	69	Based on fractals ($n = 3$): fractal dimension (FD), abundance, and lacunarity; Based on grey-level co-occurrence matrices[a] ($n = 44$): autocorrelation, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference entropy, dissimilarity, energy, entropy, homogeneity 1, homogeneity 2, informational measure of correlation 1 (IMC1), IMC2, inverse difference moment normalized (IDMN), inverse difference normalized (IDN), inverse variance, maximum probability, sum average, sum entropy, sum variance, and variance; Based on grey-level run-length matrices[a] ($n = 22$): grey-level non-uniformity (GLN), high-grey-level run emphasis (HGLRE), long-run emphasis (LRE), long-run high-grey-level emphasis (LRHGLE), long-run low-grey-level emphasis (LRLGLE), low-grey-level run emphasis (LGLRE), run length non-uniformity (RLN), run percentage (RP), short-run emphasis (SRE), short-run high-grey-level emphasis (SRHGLE), and short-run low-grey-level emphasis (SRLGLE)	[11, 15, 16]

[a]Two types of SUV discretization were used, 64 grey-level bins or a fixed bin size of 0.25 g/ml

NTR3508). Informed consent was obtained from all individual participants included in the study. Patients were included if they were 18 years or older, were diagnosed with stage IIIB or IV of NSCLC, had at least one lesion with a diameter larger than 3 cm, and were able to remain supine for 60 min during acquisition. Patients were excluded if they were pregnant or lactating, had chemotherapy in the past 4 weeks, metal implants, a body weight of more than 100 kg, or known diabetes mellitus type I or II.

Acquisition, Reconstruction, and Post-Processing

A static whole-body emission scan was started 1 h (61 ± 2 min) after injection of [^{18}F]FDG (263 ± 61 MBq). Prior to this emission scan, a low-dose CT scan (120 kVp, 50 mAs) was acquired during normal breathing. All PET data were normalized and corrected for scatter and random events, dead time, attenuation, and decay and reconstructed twice using vendor-recommended reconstruction settings. All reconstruction settings utilize a blob-based TOF list-mode-ordered subset expectation maximization algorithm with 3 iterations and 33 subsets [29]. The first reconstruction setting applied an additional Gaussian filter in order to comply with the EANM guidelines for tumor PET imaging [28]. The second reconstruction setting applied an additional post-reconstruction resolution recovery method, i.e., a maximum likelihood expectation maximization deconvolution [30] that uses the spatially variant point spread function (PSF) of the system, as implemented by the PET/CT vendor. All resulting PET images have a matrix size of 144 × 144 voxels with a voxel size of 4 × 4 × 4 mm. After reconstruction, PET image data were expressed in SUV by normalizing voxel radioactivity concentrations [kBq · ml^{-1}] to injected dose of [^{18}F]FDG [MBq] and the patient's body weight (kg). All CT images have a matrix size of 512 × 512 voxels with a voxel size of 1.2 × 1.2 × 5 mm and were rescaled to the dimensions of the PET images prior to delineation. In this way, voxel tissue fraction effects within the delineations are avoided and calculations are performed using the original non-rebinned PET images, as recommended by Uniform Protocols for Imaging in Clinical Trials (UPICT) working group [31].

Delineation

Nineteen VOIs were delineated for lesions larger than 10 ml on both PET and low-dose CT images. CT-based VOIs were drawn manually upon consensus between an experienced physician, a physician in training, and a medical physic expert, using the medical history and previously acquired contrast-enhanced CT images as prior knowledge. PET-based VOIs were drawn semi-automatically by using an isocontour method that applies a threshold of 50 % of the 3D peak SUV (SUV_{peak}, obtained using a sphere of 12-mm diameter) corrected for local background [12]. PET-based VOIs were drawn twice, i.e., both on PSF-based and EANM-compliant images.

Radiomic Features

For each VOI, 105 radiomic features were determined. These features can be divided into the following three groups (Table 2): intensity ($n = 27$), shape ($n = 9$), and texture ($n = 69$). The textural features were based on fractals, grey-level co-occurrence matrices (GLCMs), or grey-level run-length matrices (GLRMs). Features derived from GLCM and GLRM were calculated by averaging the obtained values over 13 symmetric directions in three dimensions [11]. For those features that require SUV discretization (i.e., resampling of the image intensity values), two types of discretization were used [21], 64 grey-level bins [14, 18] or a fixed bin size of 0.25 g/ml [21]. A fixed bin size of 0.25 g/ml represents the mean SUV_{max} for all 19 lesions (18 and 14 g/ml when obtained from PSF-based and EANM-compliant images, respectively) divided by 64 bins.

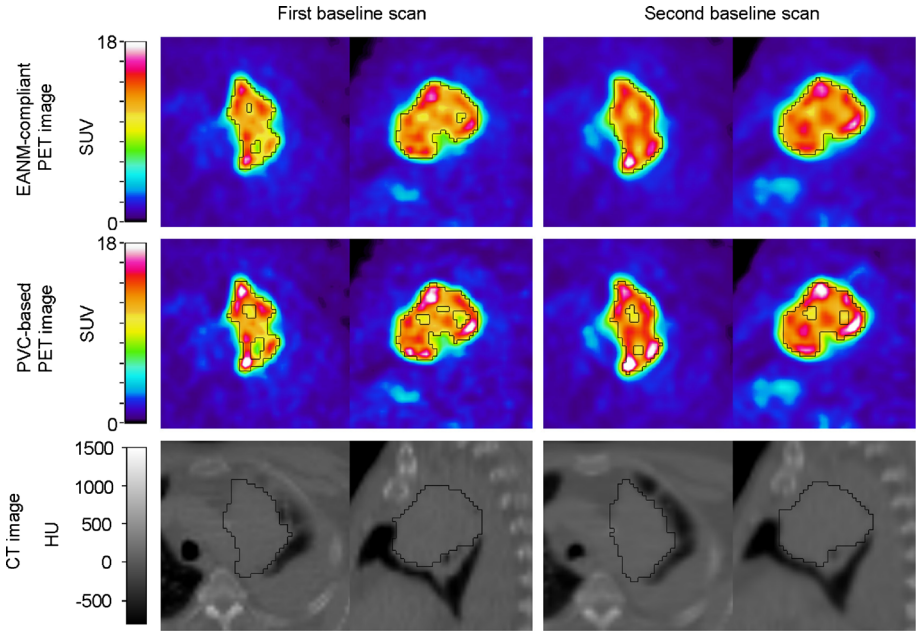

Fig. 1 Axial (*left*) and sagittal (*right*) PET/CT images of a typical NSCLC patient with (visually) rather heterogeneous [^{18}F] FDG uptake in the primary lung tumour. The *black contours* illustrate the various (CT- or PET-based) delineations. Rigid co-registration was applied for illustration purposes only to co-register the second baseline scan onto the first baseline scan using VINCI v4.23 (Max-Planck-Institute for Neurological Research, Cologne, Germany) (Color figure online).

Statistics

To assess the level of repeatability, mean relative test-retest variability (TRT_r, in %) was calculated for all 105 radiomic features by Eq. (1).

$$TRT_r = \frac{1}{n} \times \sum_{i=1}^{n} \frac{\text{test-retest}}{\text{mean(test, retest)}} \times 100\% \quad (1)$$

where n is the number of lesions. In addition, mean absolute TRT (TRT_a, in %) was calculated by $TRT_a = |TRT_r|$. A TRT closer to zero indicates a higher level of repeatability. In addition, intraclass correlation coefficients (ICCs) were calculated between the values obtained from first and second baseline scans using a one-way random single-measure model (Real Statistics Resource Pack release 3.5; www.real-statistics.com). ICC does not only take the variance within subjects but also variance between subjects into account. An ICC of 1 indicates perfect reliability. For both TRT_r and ICC, 95 % confidence intervals were calculated. A related-sample Wilcoxon signed-rank test was applied to ICC, TRT_r, and TRT_a of all features to assess whether a change in reconstruction setting or delineation significantly changed ICC, TRT_r, or TRT_a. P values less than 0.05 were considered significant.

A threshold of 0.90 for ICC was used to group features into sets of features showing an overall good, variable, or overall poor performance. This threshold is in line with the ICC found in literature for SUV_{max} [11, 13, 14]. An overall good performance means that all four possible combinations of delineation and reconstruction algorithm resulted in an ICC ≥ 0.90, whereas a variable performance means that at least one but not all combinations resulted in an ICC ≥ 0.90. An overall poor performance indicates that all combinations resulted in an ICC < 0.90. Features were considered to be sensitive to an applied delineation and/or selected reconstruction algorithm when the absolute change in ICC was at least 0.03. For these features, the best performing delineation and/or reconstruction algorithm was determined.

Results

Most intensity-, shape-, and texture-based features (98 %) have a repeatability that is comparable to those seen for simple SUV measures in literature (e.g., SUV_{max}, SUV_{mean}, and SUV_{peak}) (Supplemental Tables 1 to 12). When compared to the ICC of SUV_{mean} observed in this study, 37 % of the features showed an equal or better ICC for at least one combination of delineation and reconstruction, while 12 % of the features showed an equal or better ICC independent of delineation and reconstruction. Figure 1 shows a typical example where the various reconstruction settings and image types (e.g., functional or anatomical) resulted in different VOI. A small but significant improvement in median ICC was observed for features extracted using CT-based delineation compared to those extracted using PET-based delineation independent of the applied reconstruction setting (from 0.960 to 0.962 and from 0.953 to 0.962 for EANM-compliant and PSF-based images, respectively; Fig. 2). This is also reflected in a decrease in the number of outliers and extreme cases (Table 3), derived from the box plots (Fig. 2). In addition, a small but significant improvement in median ICC was observed for

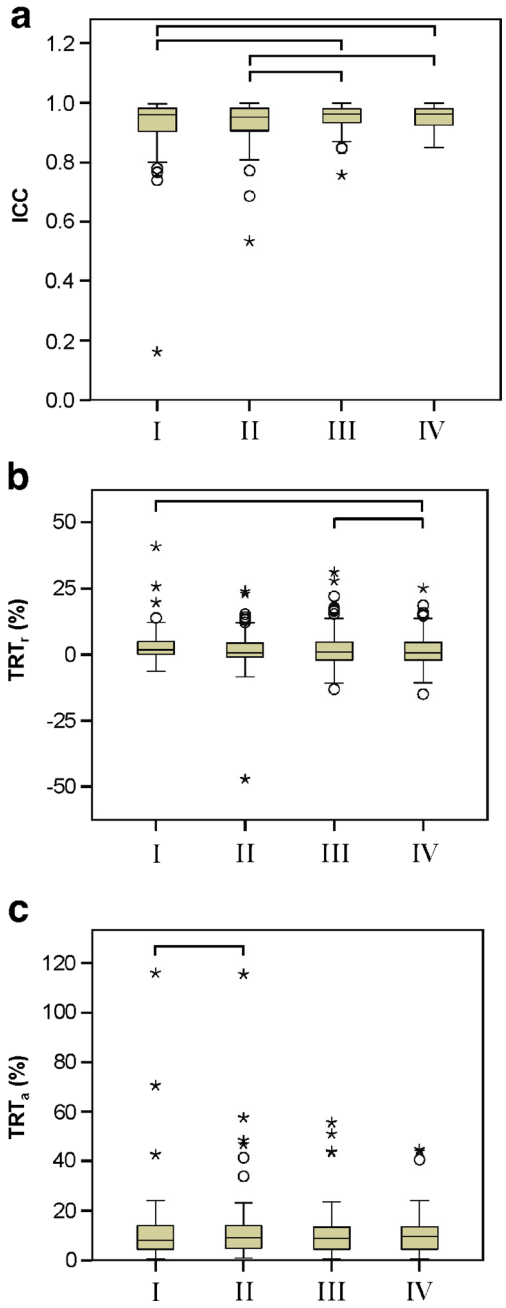

Fig. 2 Box plots of **a** ICC, **b** TRT, and **c** TRT of radiomic features extracted from EANM-compliant reconstruction with (I) PET-based or (III) CT-based delineation or PSF-based reconstruction using (II) PET-based or (IV) CT-based delineation. *Circles* illustrate outliers, and *stars* illustrate extreme cases. A *bar* indicates a statistically significant difference (p value < 0.05).

features extracted using EANM-compliant reconstruction with CT-based delineation compared to those extracted using PSF-based reconstruction with PET-based delineation (from 0.953 to 0.962). All other differences in median ICC were insignificant.

Sixty three out of 105 radiomic features showed a good performance (i.e., ICC ≥ 0.9) independent of the applied delineation or selected reconstruction algorithm, while 40 features only showed a good performance for certain combinations of reconstruction algorithm and/or delineation (Fig. 3a). More features were sensitive to a change in delineation than to a change in reconstruction (25 and 3 features, respectively), and 25 features were sensitive to a change in both reconstruction and delineation. Only fractal dimension and homogeneity 2 (obtained using 64 grey-level bins) showed an overall poor performance. After excluding these two features, the majority of the features showed less than 0.03 difference in ICC for either applied delineation and/or reconstruction (49 %; Fig. 4a). The best performance was seen using CT-based delineation (32 %), followed by EANM-compliant reconstruction or PET-based delineation (both 17 %), and PSF-based reconstruction (10 %).

More than two thirds of the intensity-based features (70 %) and one third of shape-based and texture-based features show an overall good performance (56 and 57 %, respectively; Fig. 3b). After excluding the features with an overall poor performance, most intensity-based features had a less than 0.03 difference in ICC for either applied delineation and/or reconstruction (70 %). Most shape-based features showed the best performance using PET-based delineation (56 %), while most texture-based features showed the best performance using CT-based delineation (39 %; Fig. 4b).

The percentages of both GLCM-based and GLRM-based features showing an overall good performance increased when fixed bins were applied compared to 64 grey-level bins (55 and 100 % vs 36 and 63 %, respectively; Fig. 3c, d). After excluding those features showing an overall poor performance, most features showed less than 0.03 difference in ICC for either applied delineation and/or reconstruction, except for GLCM-based features when 64 grey-level bins were applied, showing the best performance using CT-based delineation (62 %; Fig. 4c).

Discussion

The present study shows that the majority of radiomic features show a high level of repeatability that is similar or better compared to simple SUV measures such as SUV_{mean} in terms of ICC, TRT_r, and TRT_a [12, 32]. These results are in line with three previous studies by Leijenaar et al. [11], Tixier et al. [14], and Van Velden et al. [13], investigating the repeatability of various radiomic features in NSCLC patients, esophageal cancer patients, and patients with colorectal liver metastases, respectively. Data presented in these studies and the present study enable a preselection of well-performing features per category in order to further assess them for their clinical applicability.

To the best of our knowledge, this is the first study that investigates the impact of various reconstructions and delineations on the repeatability of several radiomic features, including intensity-, shape-, and texture-based features. However, this is not the first study that investigates the

Table 3. Outliers and extreme cases of radiomic features extracted from EANM-compliant or PSF-based reconstructed PET images using PET-based or CT-based delineation

Delineation	Reconstruction	ICC		TRT_r		TRT_a	
		Outliers	Extreme cases	Outliers	Extreme cases	Outliers	Extreme cases
PET-based	EANM-compliant	AUC	FD and homogeneities 1 and 2 (64B)	Variance	Cluster shade (64B and FB) and cluster prominence (64B)	–	Cluster prominence (64B and FB) and cluster shade (FB)
	PSF-based	FD and contrast (64B)	AUC	Variance, variance (FB), sum variance (FB), and autocorrelation (FB)	Skewness, cluster shade (FB), and cluster tendency (FB)	Skewness and cluster prominence (FB)	Cluster shade (64B and FB) and correlation (64B and FB)
CT-based	EANM-compliant	SRE and compactness A	AUC	Skewness, cluster shade (64B), autocorrelation (FB), cluster tendency (FB), contrast (FB), and sum variance (FB)	Cluster shade (FB) and cluster prominence (FB)	–	Skewness, cluster prominence (FB), and cluster shade (64B and FB)
	PSF-based	–	–	Skewness, autocorrelation (FB), contrast (FB), and sum variance (FB)	Cluster shade (FB) and cluster prominence (FB)	Cluster prominence (FB)	Skewness and cluster shade (FB)

Two types of SUV discretization were used, 64 grey-level bins (64B) or a fixed bin size of 0.25 g/ml (FB)

impact of reconstruction and delineation on radiomic features. A previous study by Hatt et al. [26] investigated the impact of reconstruction-based partial volume correction and various PET-based delineation on radiomic features in terms of therapy response prediction for esophageal cancer patients, showing that the performance of radiomic features were more dependent on delineation than on partial volume correction (i.e., reconstruction settings). Two studies [25, 27] investigated the effects of different reconstruction settings on the values obtained from various texture-based features. Galavis et al. [25] found that most features (80 %) showed a large variation between values (>30 %) when reconstruction settings were varied. Yan et al. [27] showed that 5 to 56 % of the features showed a large variation between values (>20 %) when reconstruction settings were varied and that zone percentage, cluster shade, and skewness should be used with caution. The level of features sensitive to the reconstruction settings is expected to be different in the present study, as the present study does not investigate differences between values obtained from features extracted from PSF-based and EANM-compliant reconstructed images but investigates whether or not they show repeatable results. Note that thresholds used in this study are arbitrary and only intended to illustrate which features are sensitive to delineation and/or reconstruction. Nonetheless, our study confirmed that many texture-based features (36 %) were sensitive to the selected reconstruction algorithm by showing a change in repeatability (i.e., showing a more than 3 % difference in ICC). In addition, we observed a large variation in repeatability for skewness and cluster shade when reconstruction settings were varied.

Recently, Leijenaar et al. [21] investigated the effects of SUV discretizations on radiomic features and concluded that the manner of SUV discretization (i.e., fixed bin size in units of SUV or a fixed number of bins) had a crucial impact on the values of various texture-based radiomic features and the interpretation thereof. They suggest that using a fixed bin size in units of SUV is more appropriate in a clinical response monitoring setting as it can incorporate changes in SUV due to a course of treatment. Our present study shows that using a fixed bin size in units of SUV results in texture-based features that show a better repeatability and a lower sensitivity to a change in delineation and/or reconstruction compared to using a fixed number of bins. A previous study [14] showed that 64 grey-level bins are best suited for extraction of radiomic features when a fixed number of bins is applied. This would, on average, translate to 0.25 g/ml for the lesions in the present study. However, a fixed bin size of 0.5 g/ml has been applied in a previous publication [11], but no further motivation is provided. A clinical study that includes outcome measures is required to validate which fixed bin size is optimal in a clinical setting. Nevertheless, this study confirms that, if a fixed bin size is best suited for clinical response monitoring, a standardized methodology in texture analysis is needed to compare results in a multicenter setting, i.e., by standardization of reconstruction settings, delineations, and SUV discretization [18, 21, 33].

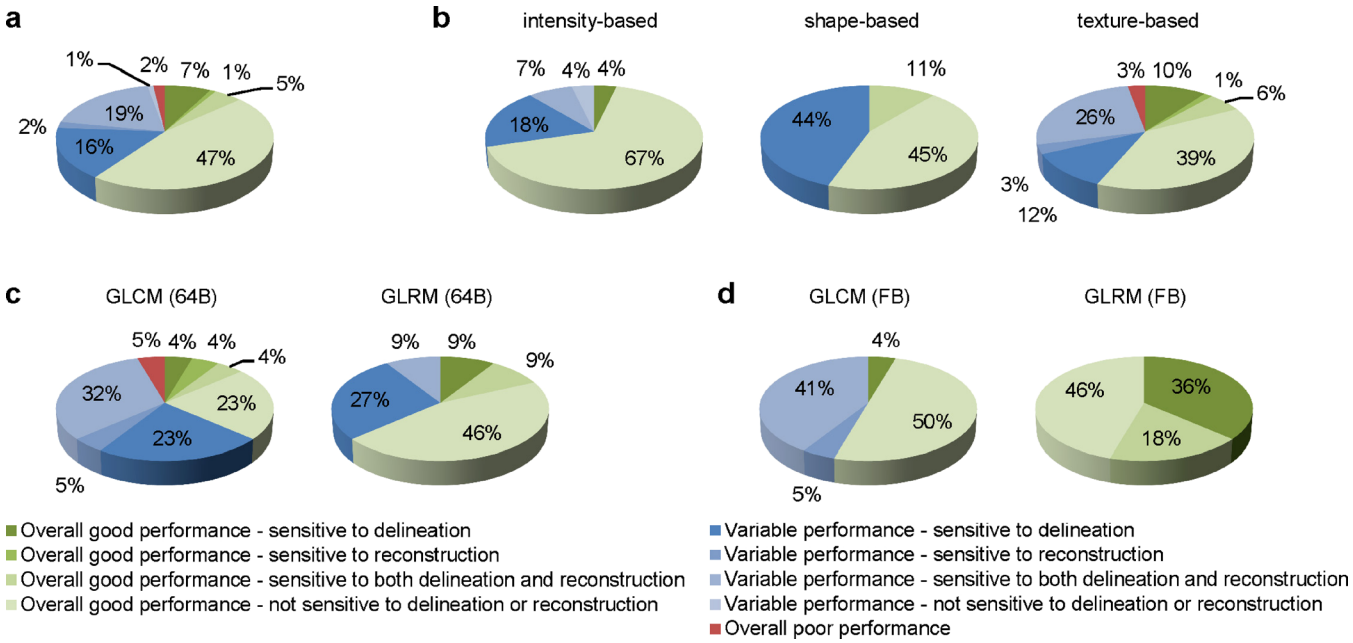

Fig. 3 Performance of radiomic features extracted from EANM-compliant or PSF-based reconstructed PET images using PET-based or CT-based delineation. Performance is given for **a** all features; **b** intensity-based, shape-based, and texture-based features; **c** GLCM-based and GLRM-based features using 64 grey-level bins; and **d** GLCM-based and GLRM-based features using fixed bins.

A limitation of this study is that the CT-based delineation is obtained manually. Therefore, these results may to a small extent be affected by inter-observer variability [11, 34]. Ideally, the effects of inter-observer variability on our results should be assessed by manual CT delineation by three observers. In this study, we aimed to minimize the impact of inter-observer variability by achieving consensus by means of discussion between three experienced observers.

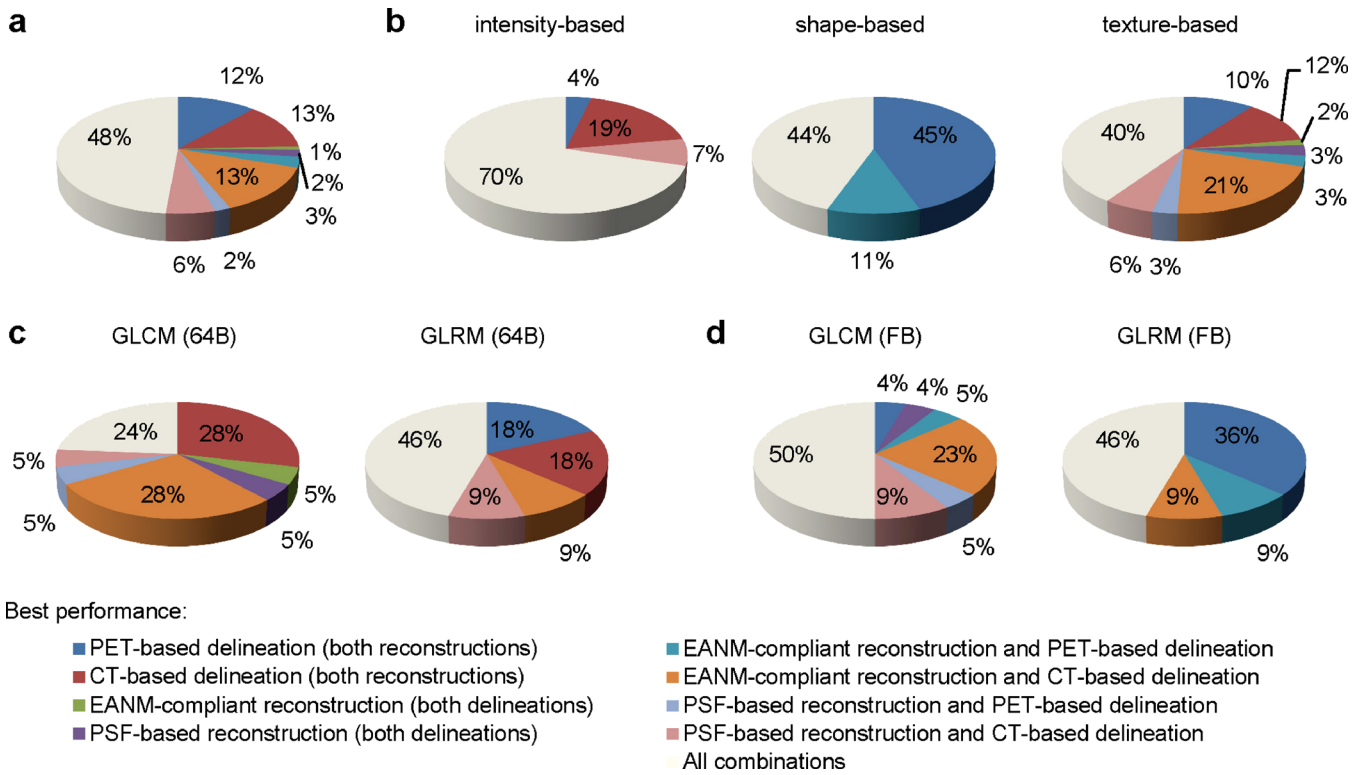

Fig. 4 Combinations of delineation and reconstruction showing the best performance, given for **a** all features; **b** intensity-based, shape-based, and texture-based features; **c** GLCM-based and GLRM-based features using 64 grey-level bins; and **d** GLCM-based and GLRM-based features using fixed bins. Features that showed a poor performance were not included.

Conclusion

In this paper, we report on the repeatability of radiomic features for NSCLC [^{18}F]FDG-PET/CT studies, showing that many features have similar TRT and ICC performance as more commonly used PET parameters, such as SUV_{max}, SUV_{mean}, and SUV_{peak}. Furthermore, PSF-based reconstructions do not necessarily result in improved repeatability of radiomic features when compared to EANM-compliant reconstructions. Performance of radiomic features depended more on delineation method than on the applied reconstruction algorithm. CT-based delineation showed favorable repeatabilities and ICCs for most radiomic features, except for shape-based features for which PET-based delineation resulted in better performance in terms of TRT and ICC.

Acknowledgments. The authors would like to thank Prof. Dr. EFI Comans for his assistance with manually delineating VOIs on CT images and NJ Hoetjes, MSc, for her assistance with delineating VOIs on PET images. This study was financially supported by CTMM, the Center for Translational Molecular Medicine: AIRFORCE project (grant 03O-103).

Compliance with Ethical Standards

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Ferlay J, Soerjomataram I, Dikshit R et al (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136:E359–E386
2. Grootjans W, de Geus-Oei LF, Troost EG et al (2015) PET in the management of locally advanced and metastatic NSCLC. Nat Rev Clin Oncol 12:395–407
3. Lambin P, Roelofs E, Reymen B et al (2013) Rapid learning health care in oncology—an approach towards decision support systems enabling customised radiotherapy. Radiother Oncol 109:159–164
4. Usmanij EA, de Geus-Oei LF, Troost EG et al (2013) ^{18}F-FDG PET early response evaluation of locally advanced non-small cell lung cancer treated with concomitant chemoradiotherapy. J Nucl Med 54:1528–1534
5. Huang W, Fan M, Liu B et al (2014) Value of metabolic tumor volume on repeated ^{18}F-FDG PET/CT for early prediction of survival in locally advanced non-small cell lung cancer treated with concurrent chemoradiotherapy. J Nucl Med 55:1584–1590
6. Toma-Dasu I, Uhrdin J, Lazzeroni M et al (2015) Evaluating tumor response of non-small cell lung cancer patients with (18) F-fludeoxyglucose positron emission tomography: potential for treatment individualization. Int J Radiat Oncol Biol Phys 91:376–384
7. van Elmpt W, Ollers M, Dingemans AM et al (2012) Response assessment using ^{18}F-FDG PET early in the course of radiotherapy correlates with survival in advanced-stage non-small cell lung cancer. J Nucl Med 53:1514–1520
8. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA (2004) Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. J Nucl Med 45:1519–1527
9. van Velden FH, Cheebsumon P, Yaqub M et al (2011) Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoral FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging 38:1636–1647
10. da Silva EC, Silva A, de Paiva AC, Nunes RA (2008) Diagnosis of lung nodule using Morans index and Gearys coefficient in computerized tomography images. Pattern Anal Applic 11:89–99
11. Leijenaar RT, Carvalho S, Velazquez ER et al (2013) Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 52:1391–1397
12. Frings V, van Velden FH, Velasquez LM et al (2014) Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. Radiology 273:539–548
13. van Velden FH, Nissen IA, Jongsma F et al (2014) Test-retest variability of various quantitative measures to characterize tracer uptake and/or tracer uptake heterogeneity in metastasized liver for patients with colorectal carcinoma. Mol Imaging Biol 16:13–18
14. Tixier F, Hatt M, Le Rest CC et al (2012) Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. J Nucl Med 53:693–700
15. Miwa K, Inubushi M, Wagatsuma K et al (2014) FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. Eur J Radiol 83:715–719
16. Goh V, Sanghera B, Wellsted DM et al (2009) Assessment of the spatial pattern of colorectal tumor perfusion estimated at perfusion CT using two-dimensional fractal analysis. Eur Radiol 19:1358–1365
17. El Naqa I, Grigsby P, Apte A et al (2009) Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recogn 42:1162–1171
18. Hatt M, Majdoub M, Vallieres M et al (2015) ^{18}F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. J Nucl Med 56:38–44
19. Orlhac F, Soussan M, Maisonobe JA et al (2014) Tumor texture analysis in ^{18}F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. J Nucl Med 55:414–422
20. Cook GJ, Yip C, Siddique M et al (2013) Are pretreatment ^{18}F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med 54:19–26
21. Leijenaar RT, Nalbantov G, Carvalho S et al (2015) The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep 5:11075
22. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48:441–446
23. Cook GJ, Siddique M, Taylor BP et al (2014) Radiomics in PET: principles and applications. Clinical Transl Imaging 2:269–276
24. Asselin MC, O'Connor JP, Boellaard R et al (2012) Quantifying heterogeneity in human tumors using MRI and PET. Eur J Cancer 48:447–455
25. Galavis PE, Hollensen C, Jallow N et al (2010) Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol 49:1012–1016
26. Hatt M, Tixier F, Le Rest CC et al (2013) Robustness of intratumor (18)F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging 40:1662–1671
27. Yan J, Lim JC, Loi HY et al (2015) Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med 56:1667–1673
28. Boellaard R, Delgado-Bolton R, Oyen WJ et al (2015) FDG PET/CT: EANM procedure guidelines for tumor imaging: version 2.0. Eur J Nucl Med Mol Imaging 42:328–354
29. Surti S, Kuhn A, Werner ME et al (2007) Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. J Nucl Med 48:471–480
30. Hoetjes NJ, van Velden FH, Hoekstra OS et al (2010) Partial volume correction strategies for quantitative FDG PET in oncology. Eur J Nucl Med Mol Imaging 37:1679–1687
31. Graham MM, Wahl RL, Hoffman JM et al (2015) Summary of the UPICT protocol for ^{18}F-FDG PET/CT imaging in oncology clinical trials. J Nucl Med 56:955–961
32. de Langen AJ, Vincent A, Velasquez LM et al (2012) Repeatability of 18F-FDG uptake measurements in tumors: a metaanalysis. J Nucl Med 53:701–708
33. Nyflot MJ, Yang F, Byrd D et al (2015) Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. J Med Imaging 2:041002
34. Persson GF, Nygaard DE, Hollensen C et al (2012) Interobserver delineation variation in lung tumor stereotactic body radiotherapy. Br J Radiol 85:e654–e660